Bandolier 151

Independent evidence-based thinking about health care

On uncertainty

When Bandolier is prised away from the computer to give a talk on EBM or similar topic, audiences are frequently frustrated at the message that there is much to distrust, and no single source of completely trustworthy information, knowledge, or wisdom. That includes Bandolier, by the way, which has no pretension to omniscience.

Audiences are aware of the complexities, because they live with them every day. They are surprised that the pointy-heads haven't solved them, and amazed when they learn that the pointy-heads have their own worries, writing their own pithy, elegant, articles, and are not at all concerned with the difficulties of day-to-day clinical practice.

Caveat lector

Caveat lector was the rule in the early days of Bandolier, and remains the rule today. There ***is*** much to distrust still, and so this issue of Bandolier is a quick reality check with examples of how much there is to mislead us, and why we have to learn and re-learn these hard lessons.

The lessons chosen include using statistical outputs rather than looking at the data, as well as examples of how a considerable literature can exist on a topic, only for it to be just plain wrong, or even possibly deliberately misleading, and examples of how, especially with adverse events, you don't find if you don't look.

But all is not lost. There are also examples of how looking for good evidence can help with design of systems that can accommodate individual patients, and help frame policy.

Bandolier shares with its readers and listeners the hope that there must be some deep and meaningful reason behind it all, but it is difficult to find sometimes. In the meantime we have to keep honing our skills at nonsense detection to avoid being bamboozled.

In this issue

NRT REVISITED

One of the interesting aspects of looking at evidence is the effect of time. Initial results for new interventions or observations often look good, but then look less good as more evidence comes in. That is why a degree of caution is sensible. Time can also affect results in the sense that those results may look better or worse as duration of treatment or observation lengthens. Clinical trials have time limits, while clinical practice may have different time horizons.

It behoves us, then, occasionally to revisit the evidence, especially when time is a feature. This has been done with nicotine replacement therapy [1], reminding us that our judgements on efficacy and effectiveness can change with time.

Systematic review

This systematic review of nicotine replacement therapy (NRT) for smoking cessation looked for randomised trials with endpoints beyond 12 months after the start of treatment. It had a terrific search strategy. The aim of trials was permanent cessation of smoking, and required that active and control arms differed only by use of NRT, so allowing supportive advice or counselling at various intensities. Use of any NRT product for any duration was allowed. Information on cessation rates at 12 months and at the longest time afterwards was abstracted.

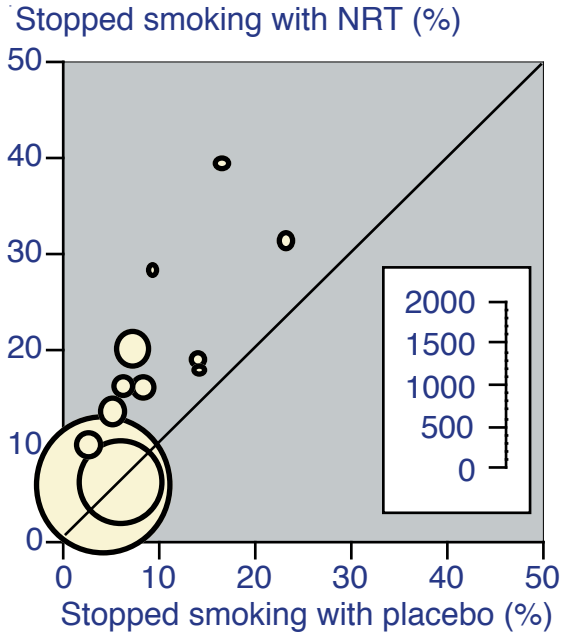
Results

There were 12 trials with 4,792 patients reporting cessation results at 2-8 years, with a weighted mean of 4.3 years. Trials used nicotine patch, gum, or spray for 3 to 52 weeks, with most using NRT for 12 weeks or longer (weighted mean 22 weeks). All but one of the trials assessed smoking cessation by breath carbon monoxide or urine cotinine measurement. Trials predominantly excluded light smokers of fewer than 10-15 cigarettes daily. Support and advice on cessation was of varying intensity.

Figure 1 shows smoking cessation results of the individual trials at the longest time. Most of the trials were small, with two contributing over 2,600 patients, and having low quit rates with both NRT and placebo.

Table 1 shows quit rates and derived statistics for the trials both at 12 months and at the longest duration. The proportion of patients who were non-smokers fell between 12 months and the longest duration, as previous quitters began

Figure 1: Smoking cessation rates at longest study duration with NRT and placebo



smoking again. A third of quitters had begun to smoke again (Figure 2) after both NRT and placebo. Figure 3 shows the results for 24 individual treatment arms of the 12 trials.

This adversely affected the efficacy of NRT as measured by the number needed to treat (Table 1). NRT rather than placebo would have to be used in 12 patients to induce one more patient to quit smoking at 12 months. But NRT would have to be used in 19 patients for one more to be a non-smoker after an average of 4.3 years.

Comment

NRT has clear efficacy in helping some patients stop smoking over the short term. The effectiveness of NRT is eroded by the propensity of former smokers to begin smoking again. This study showed that at least a third of quitters began to smoke again after NRT or placebo.

In the case of NRT the argument of cost effectiveness is governed by how many people stop smoking because of NRT. At 12 months, after an average of 22 weeks of NRT treatment, the answer is 1 patient in 12. But at longer follow up, it is more like 1 in 19.

The real importance of this study is not, however, about smoking cessation, but about how duration of observation can effect how we perceive a result. In this case, there is an argument that an intervention that looks just about useful after one year, is tipping towards irrelevance by four.

Figure 2: Smoking cessation rates at 12 months and at longest study duration

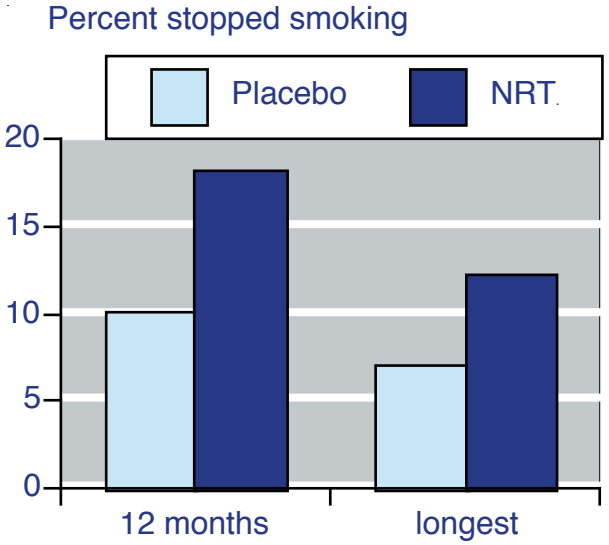
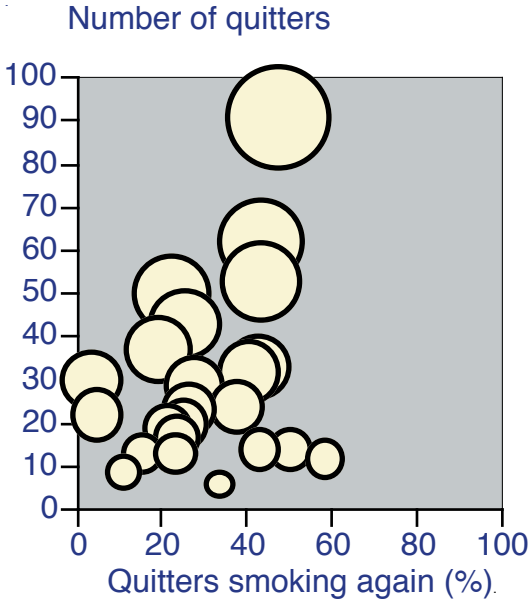


Figure 3: Percentage of quitters at 12 months smoking again at longest duration



What this does is to open something of a Pandora's box of cost effectiveness. If the effect of NRT in smoking cessation continues to diminish as recidivist quitters begin to smoke again, does the cost of the effort outweigh the health gains? It may just be easier to ban smoking in public places. In California, where bans began, smoking rates have halved.

Reference:

1 J Etter, JA Stapleton. Nicotine replacement therapy for long-term smoking cessation: a meta-analysis. Tobacco Control 2006 15: 280-285.

Table 1: Results of trials at 12 months and at longest duration

Duration of observation	Percent stopped smoking		Relative risk (95% CI)	NNT (95% CI)
	NRT	Placebo		
12 months	18.2	10.1	1.8 (1.6 to 2.1)	12 (10 to 16)
longest	12.2	7.0	1.7 (1.5 to 2.1)	19 (15 to 28)

Serious skin disorders and coxibs

Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are closely related severe acute life-threatening, drug-induced skin disorders, with mortality rates of 10-30%. The background rate in developed countries is of the order of 2 cases per million per year.

These are rare events, though, and pose a problem for calculation of event rates because individual studies tend to have few events. For instance, a study in England identified only four cases of SJS related to coxibs [1]. A study from the US [2] had 123 cases, allowing some calculations of rates.

Study

The FDA adverse event reporting database was searched for cases of SJS and TEN to March 2004 [2]. Duplicates were removed, diagnosis was confirmed, and other data examined. Cases were classified as SJS or TEN by standard criteria.

For calculation of event rates the estimated total exposure to coxibs for the first two years of marketing was estimated. The total number of dispensed prescriptions obtained from standard databases was multiplied by the average duration in days, and converted to person years of exposure.

Results

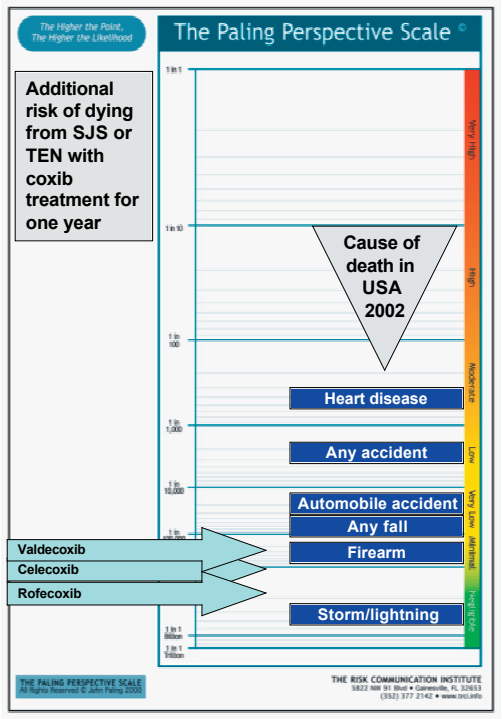
There were 123 cases in total, 63 for valdecoxib, 43 for celecoxib, and 17 for rofecoxib, predominantly SJS. For valdecoxib and celecoxib the median age was 62 years, and for rofecoxib 79 years. About 80% were women. Fourteen of these patients died.

Table 1 shows the results for valdecoxib, celecoxib, and rofecoxib for the first two years of marketing for each drug. There were more cases per million person years than the background rate of 1.9 per million per year reported from the literature, especially for valdecoxib.

Comment

There are assumptions in all these calculations, of course, not least about exposure. But this is about the best that can be done, and it can be done only because of the existence of large databases. We can go one stage further, however, and try to present the additional risk of dying because of SJS or TEN from use of coxibs.

Figure 1: Paling perspective scale for additional risk of dying from SJS or TEN



The calculation is quite simple. First we can calculate the background rate of deaths, by multiplying the background case rate of 1.9 per million person years by the death rate in this study (11%), making it 0.21 deaths per million person years. Then the final calculation is one million divided by the observed death rate minus the background death rate.

Applying this to figures in Table 1 provides rates of 1 in 313,000 for valdecoxib, 1 in 1 million for celecoxib, and 1 in 1.7 million for rofecoxib. These can be portrayed in a Paling perspective scale, as in Figure 1.

It is interesting to consider that valdecoxib was withdrawn because of these severe skin reactions, because they occurred at a rate of about 1 in 20,000 people, some 25 times the background rate. Given probable under-reporting, the true incidence was almost cetainly higher, which makes safety surveillance systems that detect a signal really important.

References:

1 D Layton et al. Serious skin reactions and selective COX-2 inhibitors. Drug Safety 2006 29: 687-696.

2 L La Grenade et al. Comparison of reporting of Stevens-Johnson syndrome and toxic epidermal necrolysis in association with selective COX-2 inhibitors. Drug Safety 2005 28:917-924.

Table 1: Results for three different coxibs

	Millions of		Number of		Rate per million person years	
Drug	Prescriptions	Person years exposure	Cases	Deaths	Cases	Deaths
Valdecoxib	16.5	1.2	58	4	49	3.4
Celecoxib	40.6	3.3	19	4	5.7	1.2
Rofecoxib	32.3	1.6	7	2	2.7	0.8

Beware statistical outputs

Suppose we have an intervention in which the NNT is 3.3 compared with placebo. That tells us that for every 3.3 people given the intervention, one more would have a treatment-specific outcome than would have occurred with placebo. Put another way, 30% more patients would have the outcome with the intervention than with placebo.

Suppose we were told the relative benefit was 3.3. That tells us that with treatment the event occurs three times more often than it does with placebo.

In both cases, if we knew the average placebo response rate, we would know how many patients would have the outcome with the intervention. Without that knowledge, the relative benefit is particularly meaningless.

Look at Figure 1. It shows the relative benefit for a series of hypothetical studies in which the placebo response rate varies between 1% and 61%, and where 30% more patients have the outcome with intervention than with placebo. The relative benefit varies from 31 at 1% placebo response to 1.5 at 61%. In all these hypothetical trials the NNT is 3.3.

From hypothetical to specific

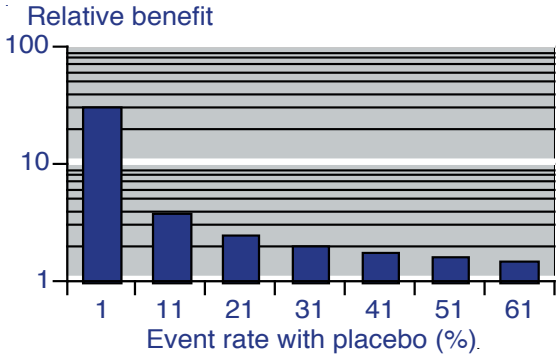
A commentary criticising the use of NNTs in acute pain [1] tries to demonstrate that different procedures produce different results, based on comparisons of relative risk and seeking aid from the dubious ally of heterogeneity tests. The results on which it is based, for paracetamol 975/1000 mg in acute pain trials, are shown in Table 1.

Relative benefit in dental and oral surgery looks higher, at about 4, than the value of below 2 seen in episiotomy and orthopaedic procedures. NNT values were the same, with overlapping confidence intervals. The additional benefit of paracetamol 975/1000 mg in these four procedures, the percentage of patients with at least half pain relief with paracetamol minus that with placebo, is also the same (Table 1), as we expect from similar NNTs.

Making decisions

What differs is what happens when we do nothing. In dental or oral surgery only 10% of patients have at least 50% pain relief, while following episiotomy or minor orthopaedic procedures 30% do so. The effect of the analgesic is the same; the procedures are different.

Figure 1: Relationship between relative risk and placebo response for 30% effect



Is paracetamol 975/1000 mg a sensible choice following episiotomy or minor orthopaedic procedures? It might be argued that it is, because most patients do well. It could also be argued that it is not a sensible choice in dental or oral procedures, because most patients do not do well.

Procedure-specific prescribing

Clearly, procedure-specific prescribing not only makes sense, but should be mandatory. Most therapy involves a package of interventions; what may be appropriate in one circumstance will not be in others.

The NNT just describes results from clinical trials. It is one way, but not the only way. Like almost all descriptors, it should not be used alone, but as a useful shorthand for busy people. In acute pain, efficacy of different analgesics in different procedures has been tested thoroughly [2]. Procedure does not affect efficacy in this particular case of clinical trials of analgesics in acute pain, and with usual provisos of sufficiency of data and limiting extrapolation.

Bandolier keeps on exhorting people to remember two important things: look at the data not just the statistics, and remember that clinical trials are not clinical practice. Using evidence does not mean putting your brain in neutral and accepting the evidence without considering appropriateness.

References:

1 A Gray et al. Predicting postoperative analgesia outcomes: NNT league tables or procedure-specific evidence? British Journal of Anaesthesia 2005 94: 710-714.

2 J Barden et al. Pain and analgesic response after third molar extraction and other postsurgical pain. Pain 2004 107: 86-90.

Table 1: Results for paracetamol 975/1000 mg in acute pain studies after different procedures

	Number of		Percent with at least half pain relief				
Procedure	Trials	Patients	Paracetamol	Placebo	Paracetamol minus placebo	Relative benefit (95%CI)	NNT (95% CI)
Dental	9	1038	37	10	**27**	3.8 (2.8 to 5.1)	3.7 (3.1 to 4.5)
Oral	6	962	37	6	**31**	4.2 (2.8 to 6.3)	3.2 (2.8 to 3.7)
Episiotomy	4	597	69	38	**31**	1.8 (1.6 to 2.2)	3.2 (2.6 to 4.2)
Orthopaedic	4	214	54	29	**25**	1.9 (1.4 to 2.6)	4.1 (2.7 to 8.6)

Policosanol for cholesterol reduction?

Policosanol is a mixture of long chain alcohols, principally from sugar cane wax, but obtainable from other plant sources. Cuban cane sugar policosanol is sold in more than 40 countries, and was described not so long ago as one of the fastest growing complementary therapy products.

In theory this wonder stuff works like plant stanols we find in margarines and other products. It is claimed to reduce lipid levels, producing major reductions in LDL cholesterol (by 25% or so) and total cholesterol (by 13%), as well as increasing HDL cholesterol by 12%.

The only fly in this particular ointment is that all the very considerable clinical trial results have come from a single source in Cuba. A single study with wheat germ policosanol showed no effect. There is a Bandolier review on this from January 2005 downloadable from the Internet site. Figure 1 shows the amazingly consistent results in total cholesterol reduction for policosanol from Cuban sugar cane, and the single trial of wheat germ policosanol. These trials used policosanol doses of 2 to 40 mg a day, mostly about 10 mg a day, over periods from one to 12 months.

What was missing was an independent trial of Cuban policosanol. We now have it [1]. Readers are invited to guess the result before reading on.

Clinical trial

The trial randomised 143 patients with high lipid levels to placebo or 10, 20, 40 and 80 mg policosanol daily for 12 weeks. The policosanol was supplied from Cuba and encapsulated in Germany. Blood samples were collected initially after a placebo run in period, and at six and 12 weeks after the start of treatment.

Figure 1: Percentage reduction in total cholesterol with policosanol in Bandolier review. Filled circle is wheat germ policosanol

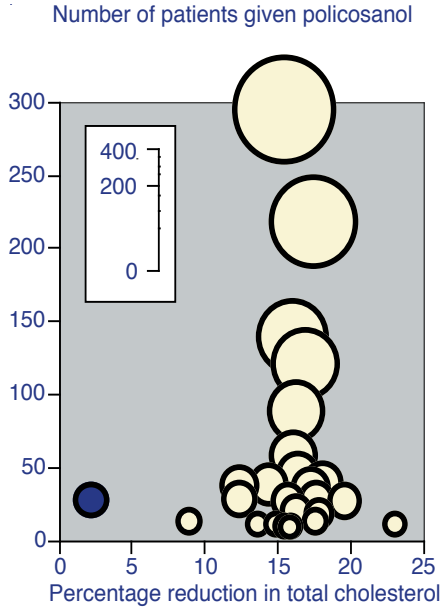
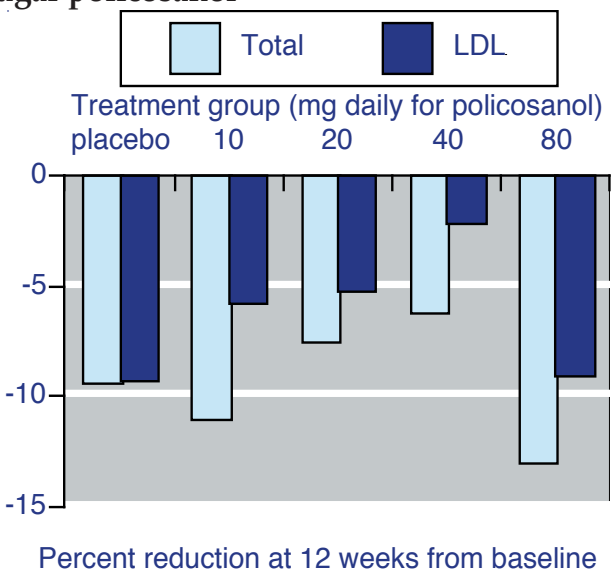


Figure 2: Percentage reduction in total and LDL cholesterol over 12 weeks with placebo and increasing daily doses of Cuban cane sugar policosanol



Results

The average age of patients was in the mid-50s, with BMI of about 27, and slightly more women than men. The trial was completed by 94% of patients, and adherence was measured at close to 100%. Initial concentrations were total cholesterol 7.3 mmol/L, LDL cholesterol 4.8 mmol/L, and HDL cholesterol 1.3 mmol/L.

There were no significant differences for any of the lipoprotein measures between baseline and the 12-week sample for any dose of policosanol or placebo, nor any difference between policosanol and placebo. Results for percentage change of 12 weeks for total and LDL cholesterol are shown in Figure 2. There was no better response at higher doses.

Comment

Now we have an independent clinical trial from outside Cuba. The result is a profound negative. Is it possible that all those Cuban trials are right and the two European ones are wrong? Or is it more likely that the trials from the group promoting policosanol, with the staggering uniformity of results shown in Figure 1, in some way overstates the case?

This is big business. Try typing policosanol into a search engine and see what you find: hundreds of thousands of sites, many trying to sell you the stuff. Any reputable regulatory agency would immediately close down any outfit trying to flog this to unsuspecting customers. And for the argument that it is safe, ask yourself whether trials with dubious efficacy measures can be trusted to report harm accurately. Think of the furore if a pharmaceutical company was allowed to market an ineffective medicine with no safety data.

Reference:

1 HK Berthold et al. Effect of Policosanol on lipid levels among patients with hypercholestrolemia or combined hyperlipidaemia. JAMA 2006 295: 2262-2269.

Checking up on serious adverse events

Suppose for a moment that you want to measure the concentration of a hormone like cortisol in blood or urine. In the absence of some of the superb immunoassays we have nowadays, you would probably argue that since cortisol is lipophilic, we should extract it from the blood or urine into an organic solvent, and measure it chromatographically.

It sounds deceptively simple. The trouble is that when examined in more detail, the efficiency of extraction of cortisol into organic solvent depends on the organic solvent used and the exact nature of what cortisol is being extracted from. So blood is different from urine, and blood samples differ from one another. Indeed, one has to be careful about the organic solvent, and some steroid hormones require freshly distilled ether, not a task undertaken lightly.

Internal standards

Some readers will be thinking that there is too much detail here, but it is necessary to introduce the concept of the internal standard. You see, if we add to our sample a minute but known amount of another steroid that is almost, but not completely, like cortisol, we can stop worrying about variability in extraction efficiency. We simply measure our cortisol concentration against the internal standard. QED.

Bandolier was intrigued to see the same concept used to evaluate the efficiency of reporting of serious adverse events in an enormous study of acupuncture in Germany [1]. It suggested huge under-reporting.

Study

The research involved the German acupuncture cohort study which began in 2001 with 12,000 doctors in private practice, predominantly general practitioners. All were certificated for Chinese acupuncture, and they provide acupuncture services to over 600,000 patients a year. The cohort involved about 200,000 patients seeking acupuncture for chronic pain – headache, back pain, arthritis – over a period of six months.

Patients usually received 10 acupuncture sessions over five weeks. Physicians reported basic data and adverse events on a standard form after the last session. They were instructed on how to complete the form, and payment by insurer depended on completing the form. Adverse events were to be counted only once, and there was no attempt to categorise adverse events by severity.

Results

Data were available for analysis on 190,924 patients receiving 1.8 million acupuncture sessions. Patients were aged from two to 97 years, with 84% aged 40 years or older. There were more females (69%) than males. Most (82%) had 10 sessions of acupuncture.

Minor adverse events were reported in 14,000 patients, with haematoma, temporary aggravation of the ailment, and vasovagal events as the most common specific minor adverse events (Table 1). Localised erythema, thought to be a desired acupuncture reaction, was reported in almost 1 in 4 patients.

By applying German population norms to the acupuncture population from five-year age and sex blocks, the expected number of deaths over the time of treatment was 180. Nine were reported (none obviously related to acupuncture), implying a rate of under-reporting of 95% (or perhaps that acupunture protects you from death?).

Serious adverse events were reported in 45 patients, including the nine deaths. This is an underlying rate of 2.4 per 10,000. In all cases the adverse event was reported to explain early termination of acupuncture sessions. Applying the correction factor for deaths, a twenty-fold higher rate of 47 per 10,000 might have been expected.

Comment

This study is not about the safety or efficacy of acupuncture, but about under-reporting of serious adverse events in clinical practice. There is an extensive literature on this, but we tend to forget about it most of the time.

In clinical trials, by contrast, there is extensive monitoring of participating doctors and patients to try to ensure complete reporting of potential harms. It is costly, but demanded for regulatory approval.

And we have not even begun to discuss causation here. Trying to figure out causation in the face of huge under-reporting smacks of the futile, and yet we agonise over it. What this study shows is that, in some cases at least, we might consider the use of an internal standard.

Table 1: Adverse events reported in 190,924 patients receiving acupuncture

Adverse event	Number	Rate per 10,000 patients	Risk
Localised erythema	46,682	2,444	1 in 4
All other minor adverse events	14,404	754	1 in 13
Specific adverse events			
Haematoma	9,896	518	1 in 19
Aggravation of ailment	2,494	131	1 in 77
Collapse, dizziness, nausea, vomiting	1,342	70	1 in 142

Reference:

1 HG Endres et al. An internal standard for verifying the accuracy of serious adverse event reporting: the example of an acupuncture study of 190,924 patients. European Journal of Medical Research 2004 9: 545-551.

How good are systematic reviews of acupuncture?

For over 150 issues now, Bandolier has been banging on about the need to examine evidence with a cold and fishy eye. If the concept of evidence-based medicine has done nothing else, it has focussed our attention on the propensity of papers in the medical sciences to give a partial view of the truth.

Bandolier looks for three things: quality, to avoid bias; validity, to make sure that what we get is what we want; size, to make sure that we aren't being pushed around by the random play of chance. Miss out on any one of these and the answer you are looking at may well be wrong.

Systematic reviews of acupuncture have tended to support the use of acupuncture, but also tended to use studies with known sources of bias, that were small, or had limited validity. What happens when you re-examine these systematic reviews using criteria of quality, validity, and size, to check out how robust their conclusions were? A systematic review of systematic reviews [1] tells us.

Systematic review

Systematic reviews published in English examining the efficacy of traditional Chinese acupuncture or mechanical or other forms of acupuncture were sought if they were published between the beginning of 1996 and the end of 2005. If there was more than one review on a particular topic, the most recent was used.

The reviews were examined for the conclusions of the original authors, and were also examined against prior criteria of quality, validity, and size. For instance, for quality, only trials that were both randomised and adequately blinded were acceptable, using sham acupuncture controls. For validity, patients had to have recognised clinical conditions, with groups comparable at baseline, and use relevant outcomes over relevant periods. For size, at least four trials and/or 200 patients were the minimum requirements.

Results

After removal of duplicate reviews, 35 systematic reviews on acupuncture remained. They examined the use of acupuncture in various painful conditions (n=18), stroke (n=2), nausea and vomiting (n=2), depression (n=2), and other conditions including insomnia, smoking cessation, weight loss, and asthma (n=11).

Most (22/35) claimed to use only randomised studies, and most were patient and assessor blind. Most (24/35) had results on fewer than 1,000 patients in total, though the number contributing to efficacy analysis was often much lower.

Of the 35 reviews, 17 concluded that there was either no evidence of benefit, or evidence of no benefit. Twelve had a qualified conclusion of some benefit for acupuncture, and six indicated that the evidence favouring acupuncture was strong (Table 1). The balance was somewhat less in favour of acupuncture for Cochrane reviews, and for those with no author affiliation with complementary medicine, but was somewhat stronger when there was some affiliation.

When criteria of quality, validity, and size were applied, none of the systematic reviews demonstrated robust evidence of effectiveness for acupuncture. After removal of poor quality studies, most reviews had only trivial amounts of good quality evidence. Only six had more than 200 patients, and in these there was no evidence of benefit.

Comment

The point here is not really about acupuncture, but about systematic reviews. Just because something is labelled as a systematic review does not mean it is any good. We have to be just as vigilant now as ever. Even a review with a Cochrane label does not make its true. Four out of 12 Cochrane reviews on acupuncture were wrong. *Caveat lector* rules, OK?

Of course, that does not stop us thinking about the efficacy of acupuncture. Large, high-quality randomised trials of acupuncture have been published since the reviews. In fibromyalgia, chemotherapy-induced nausea and vomiting, breech presentation, tension headache, and migraine, all were negative compared with sham acupuncture. One in osteoarthritis of the knee, had statistical improvement over sham acupuncture at three months, but not later. Both large trials and this review of reviews come to the same general conclusion; that over a whole range of conditions and outcomes acupuncture cannot yet be shown to be effective.

Reference:
1 CJ Derry et al. Systematic review of systematic reviews of acupuncture published 1996-2005. Clinical Medicine 2006 6:381-386.

Table 1: Support by original authors by type of review and affiliation to department of complementary therapy

	Authors' support for acupuncture		
	None	Qualified	Strong
All studies (n=35)	17	12	6
Cochrane (n=12)	8	3	1
Affiliated (n=18)	8	5	5
Not-affiliated (n=17)	9	7	1

Bleeding after tonsillectomy

Where you live in the world, and what type of health service you have, will determine whether you have tonsillectomy as a day case, or have a week in hospital. A week in hospital used to be the norm, of course, and one reason for the stay was the risk of a bleed after the operation. What determines the stay is when bleeding occurs. A new meta-analysis provides an insight [1].

Meta-analysis

A systematic review sought results of adult and paediatric tonsillectomy series that provided both the numbers of bleeding events and when they occurred. Adenoidectomies, and adenoid bleeds from adenotonsillectomies, were excluded. Bleeding events were as defined by the original authors, so included more and less serious bleeding events. The incidence of bleeding was then assessed overall for the first eight hours after operation, between eight and 24 hours, and after 24 hours.

Results

Sixteen studies provided the required data, with 27,305 patients. Studies varied in size from 94 to almost 7,000 patients. Seven included only day cases, and four were a mix of day case and longer postoperative stays. All of the studies provided information on bleeding during the first 24 hours, and 11 for periods longer than 24 hours.

In the first eight hours after operation there were 343 bleeds, a rate of 1.3%. In the period between eight and 24 hours there were 32 bleeds, a rate of 0.12%. There was a clear decay in the number of bleeds during the first 24 hours, with most frequent occurrence in the first five hours, and with only sporadic bleeding after eight hours. Beyond 24 hours there were 398 bleeds, a rate of 1.8% (Figure 1).

Comment

The results of this analysis showed that most bleeding occurred early after operation, or after 24 hours. This does not support an argument for a mandatory overnight stay after tonsillectomy. It appears that 833 patients would need to be kept in overnight to identify one case of bleeding after eight hours and before next day discharge. There will always be good reasons why some patients should have at least an overnight stay, and some of these are shown in Table 1.

Figure 1: Incidence rates of bleeding after tonsillectomy by time after the operation

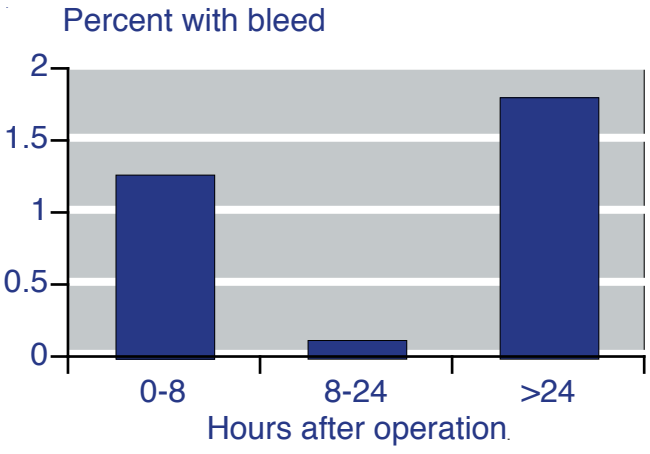


Table 1: Exclusion criteria for day case tonsillectomy for individual patients where overnight stay may be more appropriate

Exclusion criteria for day case tonsillectomy
Medical problems
Severe asthma
Diabetes
Coagulation disorder
Hypersomnia/sleep apnoea
Sickle cell disease
Epilepsy
Other conditions where overnight stay may be required
Social reasons
No access to telephone
No access to car
Only one adult at home if other children in house

The paper argues that with about 90,000 tonsillectomies a year, and with an overnight bed costing about £300, the potential saving by moving from complete overnight stays to complete day case surgery is of the order of £20-30 million for the NHS in the UK if the beds were closed or used for something else. There are also implications here about discharge before eight hours.

An interesting example, this, of how systematic review and meta-analysis can impact directly on the design and purchasing of services. It goes to the heart of what is appropriate in service design on the one hand, and what is best for individual patients on the other. Good evidence trumps dogma every time.

Reference:
1 AM Bennett et al. Meta-analysis of the timing of haemorrhage after tonsillectomy: an important factor in determining the safety of performing tonsillectomy as a day case procedure. Clinical Otolaryngology 2005 30:418-423.

EDITORS

Andrew Moore Henry McQuay
Pain Relief Unit
The Churchill, Oxford OX3 7LJ
Editorial office: 01865 226132
Editorial fax: 01865 226978
Email: andrew.moore@pru.ox.ac.uk
Internet: www.ebandolier.com
ISSN 1353-9906